国内外科学数据管理与开放共享的最新进展

张丽丽[1] 温亮明[1] 石 蕾[2] 郑晓欢[3] 黎建辉[1*]

1 中国科学院计算机网络信息中心 北京 100190
2 科技部国家科技基础条件平台中心 北京 100062
3 中国科学院 办公厅 北京 100864

摘要 作为科研活动的灵魂，科学数据既是激发科研创新的起点，也是科研活动丰富成果不可或缺的部分。国内外科学数据管理与共享从“积极温和的科学数据政策导向”和“全面细致的科学数据管理实践”两方面开展。通过对国内外发展现状的比较，笔者认为，国内的科学数据政策宏观体系尚待拓展，政策落实仍需继续积累；仍有很多学科门类需要提高数据管理意识和管理水平；科学研究的整体环境有利于开放科学数据文化的孵化，但仍需多利益相关群体的综合协调。未来，积极温和的共享趋势仍占据主流，科学数据的公、私权力争论愈演愈烈，信息技术变革与科学数据管理边界的再认识将持续推动数据开放。

关键词 科学数据管理，科学数据共享，科学数据政策，开放数据

DOI 10.16418/j.issn.1000-3045.2018.08.002

1 研究概述

无论科研领域、无论利益群体，科学数据的有效管理与开放共享使科研工作、广泛社会公众与个体普遍受益：推动科学进步，减少重复劳动并收获更多生产力，打造高效的科学政策边界[1]；推进科研与教育长期进步[2]；为社会问题带来新的解决方案[3]；缩短新产品孵化周期、满足大众信息诉求等。然而复杂的科研场景中，数据无法按照知识共同体（knowledge commons）[4]来看待和管理，数据的有效流动需要更多激励措施与质量控制、更复杂的博弈策略选择与平衡。更好地把握国内外科学数据管理与共享趋势，有利于我们识别和分析问题，对比和反思现状，以便对未来形成合理预期与研判。通过大量调研，笔者将科学数据管理与共享相关主题内容进行了整理（表1）。

*通讯作者

资助项目：国家自然科学基金项目（91546125），中国科学院院级科研项目（XXH13505-08-04），国家科技基础条件平台专项课题（2017DDJ1ZZ19）

修改稿收到日期：2018年8月12日

表 1　科学数据管理与共享相关主题

分类	热点主题	主要内容
研究对象	1. 定义与研究范围[5]	数据、信息、科学数据、数据管理、数据政策、数据权力、开放数据与开放获取等核心定义
	2. 目标和基本原则	特定数据政策的目标和原则
	3. 相关利益群体[6]	角色和责任
研究方法	4. 数据政策方法论[7]	数据政策的体系框架
		制定数据政策的方法
理论与实践	5. 数据管理[8,9]	数据资源（如属性、特征、质量）
		信息技术与标准
		全生命周期数据管理
		数据交换与交易
		数据管理的培训教育与人力资源体系保障等
	6. 数据治理[10]	数据权力[11]
		制度的逻辑框架（公共领域：条约、协议、法规、规章、法律；私有领域：非政府间的协约、合同、许可、守则）
		开放数据模型
		法律限制与开放边界
	7. 数据政策实施[12]	适用性和适用条件
		数据政策的解读
		政策实施的评价和反馈

2 国内外发展趋势研判

2.1　积极温和的科学数据政策导向

开放数据 FAIR 化提供宏观原则，数据管理计划（DMP）坚持务实操作，全方位政策体系日渐丰满，重塑科学数据开放边界从未停歇。这些积极的发展态势并非自上而下、“一刀切”的行政命令，而是与自下而上的一线科学数据生产相呼应，尤其通过技术应用、培训教育、公民科学发展与影响力全面计量等举措全面夯实。积极开放是大势所趋，温和推进则是现实所需。

2.1.1 “全面开放”FAIR化

2000 年以后，以经济合作与发展组织（OECD）、地球观测组织（GEO）、国际科技数据委员会（CODATA）等为代表的一系列国际组织推行“全面开放”（full and open）的科学数据共享政策，旨在推动科学数据资源尽可能免费、无限制性地跨界流动重用。2014 年，荷兰莱顿一场由多利益相关群体参与的名为“联合共建数据公平港口”的学术研讨会[13]提出“可发现（findable）、可访问（accessible）、可互操作（interoperable）和可重用（reusable）”的 FAIR 原则，进一步诠释现代科学数据共享的基本理念并迅速流行。FAIR 原则将科学数据资源依开放状态分为 6 类，其中“FAIR 化的元数据”“FAIR 化有限开放数据”“FAIR化开放数据”和“FFAIR 化增强版开放数据”等 4 类被认为是开放数据的主要形式。该原则在欧盟[14]、美国、澳大利亚等地普及。FAIR 化数据资产的计量研究和以“Go change，Go build，Go train”为主题的 FAIR 运动，进一步使该原则落地。

2.1.2　动态生长的科学数据开放边界

科学数据共享的深度和广度也即开放的边界。英国皇家学会研究报告《科学是一项开放的事业》[15]中明确指出，科学开放边界的制约因素包括经济利益（如数据所有权与知识产权等）、隐私权力、公共安全等。历经 4 年准备，欧盟《通用数据保护法案》（GDPR）[16]于 2016 年 4 月 14 日通过审批并于 2018 年 5 月 25 日起正式施行，旨在保护数据时代的欧洲公民免于隐私数据泄露。其核心内容确立“知情权、访问权、反对权、个人数据可携权、被遗忘权”5 种公民权利，被视为近 20 年来最重要的数据隐私规章。而数据资产确权由于科研场景的复杂性，仍值得持续探讨。可以说，生长着的科学数据开放边界正是科学数据从封闭走向开放的动态博弈。边界刻画将持续成为科学数据共享的焦点和难点。

2.1.3　全方位政策体系的日渐丰满

从组织视角来开，科学数据管理与共享的政策体系全面扩展（图 1），无论是国际与国家层面，抑或区域、领域与机构层面或者更小的单元组织。其中，领域机构层面的科学数据政策更贴近科研与数据场景，因而成为推动整个政策链条延伸与丰满的重要力量。除纵向一体化的政策体系搭建外，不同层级的政策联系也日益

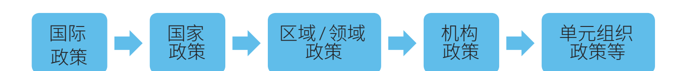


图 1　科学数据管理与共享政策体系

紧密。如以数据外交为桥梁的国际政策与国家政策的接轨，领域机构政策面向国家政策的调整与校正等。

一些组织归档了现有数据政策：美国能源部系统生物学知识库（Kbase）包括美国本土为主的生物信息学数据政策资源[17]；公平共享平台FAIRsharing[18]收录面向多领域门类112份数据政策元数据信息。欧盟与OECD合作组建国际科技政策数据库STIPCompass[19]，收集并发布包括中国在内的51个国家的科技政策，科学数据管理政策涵盖其中。

2.2 全面细致的科学数据开放管理

2.2.1 数据管理计划：从理念到实践

1995年，英国经济和社会研究委员会（ESRC）制定了数据管理计划（Data Management Plan，DMP），要求ESRC资助研究所产生的数据尽可能共享，并做好长期保存和高质量管理[20]。美国国家科学基金会（NSF）于2011年1月规定项目申请需包括数据管理计划[21]。近年来，数据管理从纸面计划逐步走向实践：关注数据类型、数据或元数据格式和内容标准、获取和共享重用政策、数据归档计划等[22]。大量图书馆、科学数据中心、科研机构、政府部门、国际与区域组织等参与了数据管理计划实践的技术支持、政策解读与培训教育。

2.2.2 新兴技术应用的持续助推

新兴技术应用助推科学数据开放共享的例子不胜枚举。以下仅就区块链推动的数据共享、公民科学激发的数据生产和数据文献倡议组织（DDI）推行的人机网络互操作等方面揭示冰山一角。

（1）区块链推动的数据共享。科学大数据全生命周期的多层次演化、流水线处理等特征[23]，对数据传输处理和共享提出全新挑战。区块链技术提供了解决方案：使用加密算法和共识机制保证安全[24]；追溯源头并“过滤”，保障数据质量；分布式决策去除中间机构，大幅提升数据共享效率[25]。医疗数据已尝试利用区块链存储共享个人健康数据[26]。此外，分布式边缘计算将发挥更大作用，通过区块链一体化快速实现数据采集、处理和分析。

（2）公民科学激发的数据生产。作为数据采集的新源头，公民科学蓬勃发展。过去22年间，生态旅行者提供的近3万张鲸鲨图片帮助科研人员有效识别了20个鲸鲨聚集点[27]。公民科学的数据价值也不容小觑。例如，公民科学联盟（Citizen Science Association，CSA）现已吸纳超过80个国家的会员注册；而该组织所参与的1 000余个重要科学计划项目，已有超百万志愿者参与其中[28]。

（3）人机网络互操作。为推动人机网络的可理解性，DDI联盟推出DDI3.3[29]，技术内容涵盖分类管理、非调查数据收集、样本和权重、问卷设计、支持DDI作为属性图、质量声明优化等，主要应用于社会学、行为科学、经济学和公共卫生领域数据的归档、发现与互操作技术指导。

2.2.3 数据出版与可信存储库

数据出版为科学数据开放管理提供新平台。以数据集及数据论文出版在近年流行，如*ESSD*（2008年）、*GigaScience*（2012年）、*Nature Scientific Data*（2015年）、《中国科学数据》（2015年）等实践。广义数据出版还包括数据存储库建设。存储库为数据集提供存储和访问平台，支持标准化的数据质量控制和完整的全生命周期管理，分为通用存储库、机构存储库、领域存储库、出版物存储库、图书馆/档案馆/博物馆以及科研项目存储库等类型[30]。可信存储库作为一种稳定可靠的数据基础设施，为包括数据出版等开放数据工作带来技术和管理资源保障。

2.2.4 繁荣的数据管理培训

数据管理培训通过实用性强的短期技能训练，指导科研实践。其中，涵盖20个国家节点的欧洲政府间组织ELIXIR[31]整体推进欧洲科学数据管理培训。英国的领域培训涉及DCC（通用）、CAiRO（艺术）[32]，DataTrain（考古学[33]、人类学）、DATUM（健康卫生）、DMT

psych（心理学）、科研数据MANTRA[34]（地学、社会科学和临床心理学）等。CODATA面向发展中国家科研人员连年提供数据管理技术培训。Data carpentry[35]由软件培训衍生而来，与世界多国合作开展培训推广。此外，数据科学专业学位教育也日渐兴盛。

2.2.5 影响力全面计量

（1）始于数据引用。2010年至今，国际科技数据委员会（CODATA）数据引用与实践工作组详细讨论了“数据引用标准与规范”[36]；2014年，美国信息科学与技术协会（ASIS&T）数据访问与保存峰会重点探讨数据引用、元数据、数据重用[37]；哈佛大学量化社会科学研究所（IQSS）在2014年启动数据引用研究项目[38]。高校图书馆与非营利性组织（如DataCite[39]、ICPSR[40]）也参与到数据引用规范的制定与推介培训中。

（2）替代计量学的社会化视角。替代计量学基于大众社交媒体、传统主流媒体、学术社交媒体、网络博客、文献管理软件等网络数据来综合评价学术成果的社会影响力（包括被浏览、保存、讨论、推荐、引用等情况）[41]。

（3）数据计量更进一步。从传统文献和参考文献扩展到数据和文献、数据和数据、数据和数据集间的多重关系，更关注“数据”“学术记录”以及“学术个人”[42]。

2.3 我国的科学数据管理与开放

2.3.1 国家科学数据政策体系概览

科学数据管理伴随着科研活动从未停歇，2000年后尤为繁荣。目前我国已形成由以政府、行业机构和领域数据中心为主体的数据政策体系（表2）。其中，《科学数据管理办法》于2018年3月17日生效。该办法首次站在国家高度、面向多领域科学数据，提出开放为主的指导原则，具有划时代意义。

此外，典型行业部门制度建设既包括数据管理办法，也涵盖政策指南，如国家海洋局《关于规范海洋生态环境监测数据管理工作的意见》（2015年2月）、交

表2　部分国内科学数据政策汇总

机构类型	发布机构	政策名称	发布时间（年/月）
政府部门	国务院	《科学数据管理办法》	2018/3
		《政务信息资源共享管理暂行办法》	2016/9
行业机构	国家海洋局	《中国极地考察数据管理办法》	2018/3
	科技部	《国家科技资源共享服务平台管理办法》	2018/2
		《国家重点基础研究发展计划资源环境领域项目数据汇交暂行办法》	2008/3
	国防科工局	《高分辨率对地观测系统重大专项卫星遥感数据管理暂行办法》	2018/1
	国家海洋信息中心	《海洋生态环境监测数据共享服务程序（试行）》	2015/12
	中国气象局	《气象信息服务管理办法》	2015/3
		《气象资料共享管理办法》	2001/11
	国土资源部	《国土资源数据管理暂行办法》	2010/9
	中国地震局	《地震科学数据共享管理办法》	2006/6
领域数据中心	国家地震科学数据共享中心	《数据共享规范标准》	2016/4
	国家生态系统观测研究共享服务平台	《国家生态系统观测研究网络数据管理与共享条例》	2013/12
	国家人口与健康科学数据共享平台	《医药卫生科学数据共享网暂行管理办法》	2009/9
		《医药卫生数据共享管理细则》	2008/1
		《中医药科研课题数据汇交管理办法》	2007/10
	国家地球系统科学数据共享平台	《地球系统科学数据共享联盟章程》	2008/12
		《地球系统科学数据共享平台章程》	2008/12

通运输部《关于推进交通运输行业数据资源开放共享的实施意见》（2016 年 9 月）等。跨部门合作共享逐步推进，如 2015 年国家林业局与国土资源部签署数据资料共享协议，建立长效共享机制。领域科学数据中心则将数据实践与数据政策并轨，值得关注。

2.3.2 相伴而生的科学数据开放共享实践

图 2 回顾了我国科学数据开放共享历程中的部分代表性事件，其中 2017 年 1月—2018 年 7 月国内科学数据共享主要实践见表 3。从所属机构看，既包括政府、科研机构，也囊括企业社会力量；从实践内容看，包括数据基础设施建设、大数据项目驱动的科学数据管理与开放、科学数据交流研讨、国际交流与合作等。相对于数据政策，数据实践先于政策并服务于政策，绝大部分领域数据政策根植于数据实践。科学数据管理先行，而开放共享尚在摸索，相关实践仍以交流研讨居多。虽也有开放数据示范平台，但广泛的数据共享实践仍有待开展。

3 比较思考与发展展望

3.1 与发达国家相比，我国科学数据政策与实践情况[43]

（1）从发展水平来看，《科学数据管理办法》新近

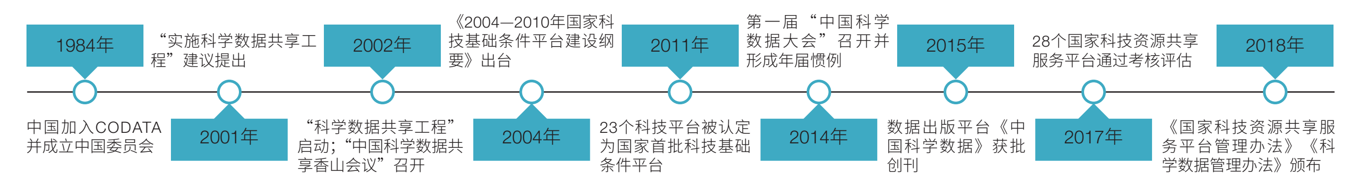


图 2 我国科学数据开放共享代表性事件

表3 2017年1月—2018年7月我国科学数据开放共享部分活动

时间（年/月）	实践主体	主要举措
2017/1	科技部、财政部	发布国家科技资源共享服务平台绩效考核与评估结果，纳入后补助的28家平台统称“国家科技资源共享服务平台”
2017/1	中国科学院国家天文台、阿里云计算有限公司	共建“天文大数据联合研究中心”
2017/4	中国科学院计算机网络信息中心	启动国家发改委促进大数据发展重大工程项目“科学大数据公共服务平台与创新应用示范”，构建科学大数据公共服务平台
2017/6	中国科学院地理科学与资源研究所、中国地理学会	召开全球变化科学研究数据出版与共享大会
2017/8	CODATA中国全国委员会、国家科技基础条件平台中心、中国科学院办公厅	第四届（2017年）中国科学数据大会，以“科学数据与国家科技计划”为主题，继续关注科学数据管理与开放共享政策、技术及基础设施建设等
2017/9	中国地震局	发布《国家地震科学数据资源手册》，梳理和汇总我国的地震科学数据资源，计划每年更新，努力推动地震领域数据共享
2017/10	国家综合地球观测数据共享平台	该平台英文网站在地球观测组织（GEO）第14届全会开通，正式面向国际范围，提供数据共享服务。
2017/10	国家科技图书文献中心	正式签署开放获取2020计划倡议《关于大规模实现学术期刊开放获取的意向书》，代表国家级文献情报机构积极参与推动学术期的开放获取
2018/1	中国科学院遥感与数字地球所等单位	中国科学院战略性先导科技专项（A类）“地球大数据科学工程”启动
2018/2	国际数字地球学会、中国科学院遥感与数字地球研究所等单位组织	全球地球科学领域首个大数据刊物——《地球大数据》创刊
2018/4	中国科学院计算机网络信息中心	“基础科学数据共享服务平台门户新版网站（www.nsdata.cn）上线，已实现超过55.83 TB数据共享，访问人次超2.3亿次，下载量超67.7 TB
2018/4	中国科学院国家天文台	LAMOST正式巡天第5年第2批光谱数据开放共享
2018/7	国际科技数据委员会中国全国委员会、国家科技基础条件平台中心、中国科学院办公厅	第五届（2018年）中国科学数据大会，关注科学计划中的数据开放政策、技术应用、知识产权保护等

颁布，贯彻落实仍需多年探索积累。基于现有的科学体系和数据资源量，数据政策宏观管理体系尚待扩展。

（2）从发展广度来看，国内的科学数据管理典型实践多集中于自然与工程科学，虽不乏社会科学数据实践典范，如国家统计数据、研究机构调查数据平台（如中国人民大学中国调查与数据中心等）等实践，但与学科科研活动规模相比，仍有很大进步空间。科学数据开放共享整体水平仍需提升，分散于研究个体手中所形成的数据黑洞仍客观存在。

（3）从发展动因来看，科学数据共享为大势所趋，但与之匹配的数据共享计量评价与激励尚不成熟。数据共享工作的动力多来自自发性或者行政约束。如何更好地将有形和无形力量结合，调动全生命周期利益相关者共同参与，关乎科研数据管理事业的未来。

3.2 未来科学数据管理与共享主流趋势

（1）积极温和的主流共享趋势仍将持续。从开放科学（open science）到开放获取（open access）再到开放数据（open data）乃至 FAIR 化实践，开放数据面向不同科研场景仍需灵活调整。例如，代表全世界 33 个国家 282 个临床研究人员的实验数据公平性调查国际委员会（ICIFTDS）组织提出，反对 14 份医学杂志关于免费共享临床试验数据的出版政策，并认为出版后 6 个月内开放数据不切实际[44]。可见，开放数据非一夜之功，而更如春雨润物，积极而温和的共享策略仍将占据主流。

（2）科学数据私权和公权博弈愈演愈烈。科学数据既应开放共享使社会公众受益，又需保护特定对象利益免于受侵。为此，科学数据确权至关重要。如何有效寻求公私权力的平衡，既需法律制度的智慧，也需信息技术扶持（如尽可能细粒度地分享数据的同时，降低对号入座的隐私侵犯风险方面的技术探索），还包括全社会的理解与参与，共享文化营建等。

（3）不容小觑的信息技术变革。信息与通信技术引领我们进入全新的数据时代并作用于科研数据资产。科学数据开放共享离不开技术支持，也对信息技术不断提出新挑战，如区块链技术的应用、公民科学的繁荣等。以开放心态迎接新技术应用，是推动开放数据管理走向成熟的新利器。

（4）科学数据管理的再认识。成熟的科学数据管理，不仅仅是和数据打交道，更涉及多方利益相关群体的参与。高效的科学数据管理活动需要科学数据管理职责细分，如机构数据资产专业细化、落实责任到人，才有可能保障数据管理达到预期。推动科学数据有效管理的基础包括但不限于机构宏观数据管理职能、数据治理机构（制度制定者与践行者）、团队文化以及成果度量评价等[45]。

4 结语

综上，通过文献调研与交流实践，总结了国内外科学数据管理与共享研究与实践的主要进展。基于国内外实践对比，从发展水平、发展广度和动因三方面出发，认为国内的科学数据管理领域实践日趋成熟，但宏观发展仍需更多积累，不同学科领域间数据管理水平仍存在显著差异，信息技术的灵活运用与科学数据管理边界的拓展将是提升科学数据管理发展的重要推动力量等。

致谢 感谢国家留学基金委员会对本文第一作者在美访学研究期间的支持。

参考文献

1 Pfenninger S, De Carolis J, Hirth L, et al. The importance of open data and software: Is energy research lagging behind? Energy Policy, 2017, 101: 211-215.

2 NSF National Science Board. Long-lived digital data collections: enabling research and education in the 21st century. Washington DC: NSF, 2005.

3 Schrier B. Government open data: Benefits, strategies, and use. The Evans School Review, 2014, 4(1): 12-27.

4 Fecher B, Friesike S, Hebing M. What drives academic data sharing? PLoS ONE, 2015, 10(2): e0118053.

5 DataOne. Sharing data: legal and policy considerations. [2018-07-02]. https://www.dataone. org/best-practices/sharing-data-legal-and-policy-considerations.

6 Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC: The National Academies Press, 2015.

7 Robert C, Zeckhauser R. The Methodology of Normative Policy Analysis. Journal of Policy Analysis and Management, 2011, 30(3): 613-643.

8 Briney K, Goben A, Zilinsk L. Do You Have an Institutional Data Policy? A review of the current landscape of library data services and institutional data policies. Journal of Librarianship and Scholarly Communication, 2015, 3(2): eP1232.

9 Naughton L, Kernohan D. Making sense of journal research data policies. Insights, 2016, 29(1): 84-89.

10 Kanous A, Brock E. Contractual Limitations on Data Sharing. [2016-08-18]. http://deepblue.lib.umich.edu/bitstream/2027.42/123016/1/ContractualLimitationsonDataSharing150411-1.pdf.

11 Defense Information System Agency. Data rights. [2018-07-01]. https://disa.mil/About/Legal-and-Regulatory/DataRights-IP/DataRights.

12 Tumwesigye B T, Nakanjako D, Wanyenze R, et al. Policy development, implementation and evaluation by the AIDS control program in uganda: A Review of the Processes. Health Res Policy, 2013, (11): 7.

13 Wilkinson M D, Dumontier M, Aalbersberg I J, et al. The FAIR Guiding Principles for scientific data management and stewardship. Nature Scientific Data, 2016, 3: 167-172.

14 European Commission Directorate-General for Research & Innovation. H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. [2018-07-01]. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

15 Boulton G, Campbell P, FREng B C, et al. Science as An Open Enterprise. London: The Royal Society Science Policy Centre, 2012.

16 EU GDPR Information Portal. GDPR Key Changes. [2018-07-03]. https://www.eugdpr.org/the-regulation.html.

17 KBase. Data policies. [2018-07-01]. https://kbase.us/data-policy-and-sources/.

18 FAIRsharing.org. Policies. [2018-07-01]. https://fairsharing.org/policies/.

19 Stipcompass. Homepage. [2018-07-01]. https://stip.oecd.org/stip.html.

20 Economic and Social Research Council. Research Data Policy. [2018-07-01]. https://esrc.ukri.org/funding/guidance-for-grant-holders/research-data-policy/.

21 National Science Foundation. Dissemination and Sharing of Research Results. [2018-07-01]. https://www.nsf.gov/bfa/dias/policy/dmp.jsp.

22 National Science Foundation. Chapter II-Proposal Preparation Instructions. [2018-07-01]. https://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp.

23 黎建辉, 沈志宏, 孟小峰. 科学大数据管理：概念、技术与系统. 计算机研究与发展, 2017, 54(2): 235-247.

24 袁勇，王飞跃. 区块链技术发展现状与展望. 自动化学报, 2016, 42(4): 481-494.

25 丁伟, 王国成, 许爱东, 等. 能源区块链的关键技术及信息安全问题研究. 中国电机工程学报, 2018, 38(4): 1026-1034.

26 薛腾飞, 傅群超, 王枞, 等. 基于区块链的医疗数据共享模型研究. 自动化学报, 2017, 43(9): 1555-1562.

27 周舟. 综述：公民科学在美国日益兴起. [2018-08-19]. http://www.xinhuanet.com/world/2018-06/14/c_1122987853.htm.

28 CSA. The power of Citizen Science. [2018-08-19]. http://www.citizenscience.org/.

29 The DDI Alliance. New DDI 3.3 specification available for public review and comment. [2018-06-30]. http://www.ddialliance.org/announcement/new-ddi-33-specification-available-for-public-

review-and-comment.

30 ICSU World Data System. Core Trustworthy Data Repositories Requirements. [2018-07-05]. https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf.

31 ELIXIR. Training services. [2018-07-01]. https://www.elixir-europe.org/services/training.

32 Jisc. CAiRO: Curating Artistic Research Output. [2018-07-01]. https://www.webarchive.org.uk/wayback/archive/20140614073310/http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain/cairo.aspx.

33 Archaeology Data Service. Data train. [2018-07-01]. http://archaeologydataservice.ac.uk/ learning/DataTrain.xhtml.

34 Mantra. Home. [2018-07-01]. https://mantra.edina.ac.uk/.

35 Data Carpentry. Home. [2018-07-02]. http://www.datacarpentry.org.

36 Codata. Open Data and Information for a Changing Planet. [2018-07-03]. http://codata2012.tw/.

37 Association for Information Science & Technology. Research Data Access & Preservation Summit. [2018-07-03]. https://www.asist.org/rdap/past-events/#comments.

38 The Institute for Quantitative Social Science. IQSS Data Science: Aiding Reproducible Research By Adding Provenance in Data Citations. [2018-07-03]. https://www.iq.harvard.edu/news/iqss-data-science-aiding-reproducible-research-adding-provenance-data-citations.

39 DataCite. Why is it so important to cite data? [2018-07-03]. https://www.datacite.org/cite-your-data.html.

40 ICPS. Find & Analyze Data. [2018-07-03]. https://www.icpsr.umich.edu/icpsrweb/ ICPSR/ curation/citations.jsp.

41 刘俏. Altmetrics指标与传统计量指标对学术论文评价的关系研究. 情报理论与实践, 2018, 41(7): 60-64.

42 顾立平. 数据级别计量——概念辨析与实践进展. 中国图书馆学报, 2015, 41(2): 56-71.

43 宋戈, 胡文静. 国外强制性开放科学数据政策调研与分析. 图书情报工作, 2016, 60(9): 61-69.

44 Murray T. Researchers oppose data-sharing proposal. Canadian Medical Association Journal, 2016, 188(14): E336

45 Anne Marie Smith. Foundations of data stewardship. [2018-07-01]. https://www.ewsolutions.com/foundations-data-stewardship/.

Progress in Scientific Data Management and Sharing

ZHANG Lili[1] WEN Liangming[1] SHI Lei[2] ZHENG Xiaohuan[3] LI Jianhui[1*]

(1 Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China;

2 National Science and Technology Infrastructure Center, Ministry of Science and Technology of the People's Republic of China, Beijing 100062, China;

3 Administrative Office, Chinese Academy of Sciences, Beijing 100864, China)

Abstract As the soul of research activities, scientific data is not only the inputs for stimulating scientific research innovation, but also the indispensable outcomes during research. Tracing the policies and practices on scientific data management and sharing, main development trends have been identified as "positive and gentle research data policy" as well as "full and delicate scientific data management". Upon comparison of domestic and international policies and practices, we figure out that the national scientific data policy development is still at its early stage

*Corresponding author

while multidisciplinary practices on research stewardship are still on demand. The open data trend could positively nourish better scientific data management and establishment of research data sharing culture. More involvement of different stakeholders is still in urgent need. Further, gentle and positive sharing trend will still last; public data rights shall continue tradeoff against private data rights fiercely. IT innovation and re-definition of scientific data management could benefit data sharing as well.

Keywords scientific data stewardship, scientific data sharing, scientific data policy, open data



张丽丽 北京大学管理学博士，中国科学院计算机网络信息中心高级工程师，国际科技数据委员会（CODATA）数据政策委员会成员，《中国科学数据》编辑部副主任。主要研究领域：科学数据管理（数据管理、数据开放），信息经济学。E-mail: zhll@cnic.cn

ZHANG Lili Received her Ph.D. degree from Peking University and she is now a senior engineer in Computer Network Information Center, Chinese Academy of Sciences (CAS). As a member of CODATA Data Policy Committee, she also serves as the Deputy Director of editorial office of *China Scientific Data* (www.csdata.org), the leading data journal publishing bilingual data papers of multidisciplinary fields in China. Her research mainly focuses on scientific data management (data stewardship, open data) and information economics. E-mail: zhll@cnic.cn



黎建辉 中国科学院计算机网络信息中心正研级高工，博士，博士生导师，国际科技数据委员会（CODATA）执委，大数据应用服务技术北京市工程实验室主任。长期致力于推动科学数据开放、共享与应用服务，负责中国科学院科学数据云服务体系构建、云服务平台研发以及数据密集型科研应用创新工作。当前主要从事大数据资源开放共享、大数据管理技术、大数据计算与分析技术等方面研究工作。E-mail: lijh@cnic.cn

LI Jianhui Professor of Computer Network Information Center, Chinese Academy of Sciences (CAS). He is also the Executive Member at CODATA, and Director of Beijing Engineering Laboratory for Big Data Application Service Technologies. Prof. Li has long been dedicated to promoting data openness, sharing, and application. His major tasks include: to construct the CAS Data Cloud Service System, to develop cloud service platforms, and to innovate data-intensive scientific applications. He is currently engaged in technologies concerning big data sharing, curation, computing, and analysis. E-mail: lijh@cnic.cn

■责任编辑：岳凌生